



ANALYSIS USING ANT COLONY OPTIMIZATION TECHNIQUES TO MINE LUNG CANCER DATA FOR THE PURPOSE OF INCREASING OR DECREASING THE DISEASE PREDICTION VALUE

K. Gokul¹, Dr. R. Sankarasubramanian²

¹ Research scholar, Department of Computer Science, Erode Arts and Science College, Erode- 09

² Associate professor, Department of Computer Science, Erode Arts and Science College, Erode-09

ABSTRACT

The leading cause of death for both men and women is cancer. The unchecked proliferation of aberrant cells that begin in one or both lung most often in the cell lining the air passageways. Small cell lung cancer and non-small cell lung cancer are the two primary forms. Non smokers Get Lung Cancer at a Rate of 10–15%. Smokers Make Up 50% of the Case. The longer someone smokes and the more cigarettes they smoke, the higher their risk of developing lung cancer. Lung cancer has become more common. Age, sex, wheezing, shortness of breath, and chest pain are among the symptoms that can indicate a patient's likelihood of developing lung cancer. Data mining algorithms, such as classification, decision tables, naïve-based, ant colony optimization, lung cancer prediction, and data mining techniques, are used to detect lung cancer disease in its early stages. According to this paper, early detection of lung cancer can completely cure the disease and help doctors save patients' lives. Ant colony optimization data mining techniques are useful for improving or decreasing the disease prediction value of lung cancer data.

KEYWORDS: ACO (Ant Colony Optimization), Classification, Data Mining, Decision Table, Lung Cancer Prediction

1. INTRODUCTION

A cancer that originates in the lungs is called lung cancer. Globally, lung cancer is the leading cause of cancer-related mortality. Lung cancer has become the most frequent cancer among men worldwide due to its significant growth in occurrence. The world's greatest significant avoidable cause of cancer is smoking. A person may get symptoms elsewhere in their body if the initial lung cancer has spread to other parts. Lung cancer frequently spreads to lymph nodes, bones, the brain, the liver, and other regions of the lungs. There is a substantial correlation between smoking cigarettes and the incidence of lung cancer; tobacco usage is linked to lung cancer in around 90% of cases. Lung cancer risk rises in proportion to the number of times. Most people know that smoking causes cancer, but may not realize how many nonsmokers get

lung cancer, too. The purpose of this work is finding the risk factor of lung cancer and classifying the smokers and non-smokers who are all caused by lung cancer by using the data mining Technique.

2. RELATED WORKS

P. Thangaraju and others, [1] Using Data Mining Techniques to Mine Lung Cancer Data for Smokers and Non-Smokers A cancer that originates in the lungs is called lung cancer. The largest risk factor for lung cancer is smoking. The risk of lung cancer increases with the number of years smoked and the amount of cigarettes smoked. Although lung cancer can strike anyone at any age, most cases occur in those between the ages of 65 and 70. Lung cancer can even strike young persons who have never smoked. This paper

aims to identify the lung cancer risk factor. It is intended to

prevent lung cancer in humans.

A model for almost detection and accurate diagnosis of the disease was proposed by Krishnaiah V. et al. [2], which will assist the physician in preserving the patient's life. It can forecast a patient's risk of developing lung cancer using generic symptoms such as age, sex, wheezing, shortness of breath, and pain in the arm, chest, or shoulder.

Several data mining and ant colony optimization strategies were proposed by ParagDeoskar et al. [3] for suitable rule generation and classification, which pilot to accurate cancer categorization. Furthermore, it offers a fundamental basis for future advancements in medical diagnostics. The characteristics of the ant colony optimization (ACO) technique are also covered in this work. The disease prediction value can be increased or decreased with the use of ant colony optimization.

T. Sowmiya and others [4] One of the most deadly forms of cancer in the world is lung cancer. The unchecked proliferation of cells in lung tissues can cause these diseases to spread throughout the world. Patients afflicted with cancer may live longer and have a better prognosis if the disease is discovered early. In this paper we survey several aspects of data mining procedures which are used for lung cancer prediction for the patients. Data mining concepts are useful in lung cancer classification. We also reviewed the aspects of ant colony optimization (ACO) technique in data mining. Ant colony optimization helps in increasing or decreasing the disease prediction value of the diseases. This case study assorted data mining and ant colony optimization techniques for appropriate rule generation and classifications on diseases, which pilot to exact Lung cancer classifications. In addition, it provides a

basic framework for further improvement in medical diagnosis of lung cancer.

The development of significant pattern prediction tools for a lung cancer prediction system was proposed by Prashant Naresh et al. [5]. The early prediction of lung cancer is expected to be crucial for both the diagnosis process and an effective preventive strategy. The lung cancer risk prediction system should be useful in identifying an individual's predisposition for lung cancer.

3. DATA MINING TECHNIQUE

The process of automatically gathering vast amounts of data with the goal of identifying hidden patterns and examining the connections between various data types in order to create prediction models is known as data mining. Two types of data analysis that can be used to generate models representing significant data classes or forecast future data trends are classification techniques and prediction. This kind of study can aid in giving us a deeper comprehension of the material overall.

4. DATA SET

For the mining algorithms to be more predicatively accurate, the dataset employed in this model needs to be more exact and accurate. Whatever is gathered can lack certain attributes or be irrelevant. To ensure that the data mining process yields the best results possible, these must be managed effectively. Age, gender, height, weight, radon gas, asbestos, and smoking habit air contamination, lung radiation therapy, HIV/AIDS, and organ transplantation.

In the method mainly decision tree is used for predicting the Lung Cancer Disease from the given data set instances and the proposed model contains three different types of decision tree algorithms such as Naïve Bayes, Decision Table and j48 are applied on type Lung Cancer Disease dataset in the WEKA tool and the performance is calculated. Here the framework can be given as below and the performance can be obtained based on the time taken to build the tree and correctly classified instances.

The duration required by the algorithms (J48, Decision Table, and Naive Bayes) to construct the decision tree within the Weka tool.

The time is expressed in milliseconds in the table above. While the decision table takes 0.05 ms and the Naive Bayes takes 0.01 ms to develop, the J48 takes 0.03 ms to build the decision tree in the Weka tool. We can conclude with ease from the above table that the Naive Bayes algorithm provides the best performance in terms of time. The 303 examples in the dataset are used as test cases for the classification methods. The examples that are accurately classified provide insight into the algorithms' performance. Every algorithm has a distinct classification. The instances which are correctly classified using the WEKA tool can be given as below.

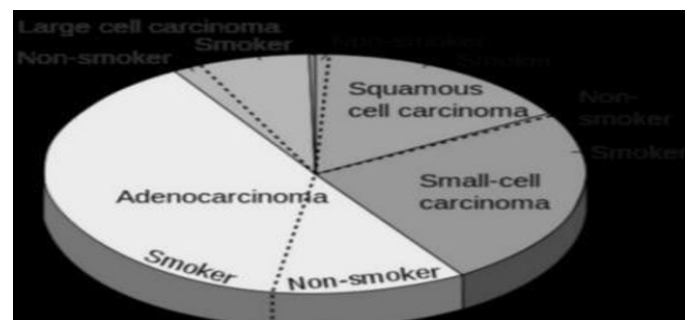
The classifier's accuracy is measured by how well it can classify unlabeled data.

$$\text{Accuracy} = \frac{\text{Number of Object Currently Classified}}{\text{Total Number of Object in the Test Set}}$$

They are frequently associated with a history Accuracy measure is the right classification of the data. With the aid of the Weka tool, the classification approach is utilized in this work to examine the risk factors for smokers and non-smokers based on each human cell and the stages of lung cancer. It will enable early treatment and identification of lung cancer problems.

5. CLASSIFICATION OF LUNG CANCER

With the use of the Weka tool, the classification approach is utilized to examine the risk factors for smokers and non-smokers based on each human cell and the stages of lung cancer.



Classification of lung cancer based on cell carcinoma.

5.1 Aden carcinoma

Aden carcinoma is a common histological form of lung cancer. Nearly 40% of lung cancers are aden carcinoma, which usually originates in peripheral lung tissue. Most cases of aden carcinoma are associated with smoking; however, among people who have smoked fewer than 100 cigarettes in their life times("never-smokers"), Aden carcinoma is the most common form of lung cancer.

5.2 Squamous cell carcinoma

The flat cells that line the inside of the lungs' airways are called squamous cells, and these Forms of smoking and are typically located in the lungs' center, close to a bronchus.

They are frequently associated with a history Accuracy measure is the right classification of the data. With the aid of the Weka tool, the classification approach is utilized in this work to examine the risk factors for smokers and non- smokers based on each human cell and the stages of lung cancer. It will enable early treatment and identification of lung cancer problems.

5. CLASSIFICATION OF LUNG CANCER

With the use of the Weka tool, the classification approach is utilized to examine the risk factors for smokers and non-smokers based on each human cell and the stages of lung cancer.

5.3 Large cell carcinoma

Ten to fifteen percent of lung cancers are of this sort. Because of its propensity for rapid growth and spread, treatment may be more difficult. big cell neuroendocrine carcinoma is a

fast- growing subtype of big cell carcinoma that has many characteristics with small cell lung cancer.

5.4 Small cell carcinoma

Small cell carcinoma often starts in the bronchi near the center of the chest, and it tends to spread widely through the body fairly early in the course of the disease.

6. PRE-DIAGNOSIS TECHNIQUES

Pre-diagnosis aids in determining or limiting the likelihood of lung cancer disease screening. Insulin resistance, alcoholism, smoking, and obesity were risk factors and symptoms that had a statistically significant impact on the pre- diagnosis stage. The diagnostic and prognostic issues with lung cancer primarily fall within the category of the much talked-about categorization issues. Numerous academics in the domains of statistics, data mining, and computational intelligence have expressed interest in these issues. While most cancer research is clinical or biological in character, data-driven statistical research is becoming frequently used as a complement. One of the most fascinating and difficult jobs where to develop data mining applications is predicting the course of an illness. Medical research groups are able to access vast amounts of medical data through the use of computers equipped with automated instruments. As a result, medical researchers are increasingly using Knowledge Discovery in Databases (KDD), which incorporates data mining techniques, as a research tool to find and take advantage of patterns and relationships among a large number of variables and to predict disease outcomes based on historical cases stored in datasets. This study aims to compile a number of reviews and technical publications about lung cancer diagnosis. It provides a summary of the research being done right now to improve lung cancer diagnosis using data mining techniques on a variety of lung cancer datasets.

Data mining Technique of prediction technique is based on systematic study of the statistical factors, symptoms and risk factors associated with Lung cancer. Non-clinical symptoms and risk factors are some of the generic indicators of cancer diseases. Initially

the parameters for the pre-diagnosis are collected by interacting with the pathological, clinical and medical oncologists (Domain experts).

7. LUNG CANCER SYMPTOMS

The following are the generic lung cancer symptoms

- Coughing up blood (heamoptysis) or bloody mucus.
- Chest, shoulder, or back pain that doesn't go away and often is made worse by deep Hoarseness
- Weight loss and loss of appetite
- Increase in volume of sputum
- Wheezing
- Shortness of breath
- Repeated respiratory infections, such as bronchitis or pneumonia
- Repeated problems with pneumonia or bronchitis
- Fatigue and weakness
- New onset of wheezing

- Swelling of the neck and face
- Clubbing of the fingers and toes. The nails appear to bulge out more than normal.
- Paraneoplastic syndromes which are caused by biologically active substances that are secreted by the tumor.
- Fever
- Hoarseness of voice
- Puffiness of face
- Loss of appetite
- Nausea and vomiting

7.1 Lung Cancer Risk Factors

Lung Cancer is affected by many Risk Factors. The Risk Factors are as follows below. Such as,

- Smoking: Beedi, Cigarette and Hookah
- Second-hand smoke
- Radon exposure
- Air pollution
- Insufficient consumption of fruits & vegetables
- Suffering with other types of malignancy.

8. EXISTING METHOD

In the United States and around the world, lung cancer is the leading cause of cancer- related fatalities in both men and women. Smoking cigarettes is the main risk factor for developing lung cancer. The degree to which lung cancer has progressed throughout the body is indicated by the stage of the disease. Overall, non- smokers account for 10–15% of lung cancer cases. (An additional 50% happen to ex- smokers). Women make up two thirds of nonsmokers with lung cancer, and 20% of these cases in women are in people who have never smoked. Lung cancer is the leading cause of death.

9. CONCLUSION

In this paper Data mining plays a major role in extracting the hidden information in the medical database. The purpose of data preprocessing is to raise the caliber of the data. This study tested the dataset, and it was completed satisfactorily. using a variety of data mining categorization methods. Data mining is thought to have a major impact on lung cancer research and ultimately enhance the standard of care for patients with lung cancer. It can also be applied with a variety of classification methods. Sometimes people with lung cancer, especially those in advanced stages of the disease, do not exhibit the typical signs of their illness. Due to a lack of knowledge, many patients were unaware of the existence of lung cancer at an early stage. The focus of this endeavor is to identify the target population for additional lung cancer screening, in order to enable the Reductions in the mortality rate and prevalence are possible. The accuracy classification techniques developed in this paper's study of multiple datasets help to increase and decrease disease, improve prediction values, improve the lack of awareness among lung cancer patients, and ultimately improve the quality of care provided to lung cancer patients.

REFERENCES

1. Thangaraju P, Karthikeyan T, Barkavi G, Mining Lung Cancer Data for Smokers and Non-Smokers by Using Data Mining

- Techniques, International Journal of Advanced Research in Computer and Communication Engineering Vol. 3, Issue 7, July 2014.
2. Krishnaiah V, Narsimha G, Subhash Chandra N , Diagnosis of Lung Cancer Prediction System Using Data Mining Classification Techniques, International Journal of Computer Science and Information Technologies, Vol. 4 (1) , 2013, 39 – 45.
 3. Parag Deoskar, Dr. Divakar Singh, Dr. Anju Singh, Mining Lung Cancer Data And Other Diseases Data using Data Mining Techniques, A Survey, Volume 4, Issue 2, March – April (2013).
 4. Sowmiya T, Gopi M, Thomas Robinson, Optimization of Lung Cancer using Modern Data Mining Techniques, International Journal of Engineering Research Volume No.3, Issue No.5.
 5. Prashant Naresh,, Early Detection of Lung Cancer Using Neural Network Techniques, Prashant Naresh Int. Journal of Engineering Research and Applications www.ijera.com ISSN : 2248-9622, Vol. 4, Issue 8(Version 4), August 2014, pp.78-83.